

OTD Observations of Continental US

Ground and Cloud Flashes

William Koshak¹

1. National Space Science & Technology Center, NASA Marshall Space Flight Center,
Huntsville, Alabama 35805, USA

ABSTRACT: Lightning optical flash parameters (e.g., radiance, area, duration, number of optical groups, and number of optical events) derived from almost five years of Optical Transient Detector (OTD) data are analyzed. Hundreds of thousands of OTD flashes occurring over the continental US are categorized according to flash type (ground or cloud flash) using US National Lightning Detection NetworkTM (NLDN) data. The statistics of the optical characteristics of the ground and cloud flashes are inter-compared on an overall basis, and as a function of ground flash polarity. A standard two-distribution hypothesis test is used to inter-compare the *population* means of a given lightning parameter for the two flash types. Given the differences in the statistics of the optical characteristics, it is suggested that statistical analyses (e.g., *Bayesian Inference*) of the space-based optical measurements might make it possible to successfully discriminate ground and cloud flashes a reasonable percentage of the time.

1. INTRODUCTION

Data collection from the OTD ended in the year 2000 after nearly 5 yrs of unprecedented global lightning observations. Calibration [Koshak et al. 2000], validation [Boccippio et al. 2000] and performance [Boccippio et al. 2002] studies of OTD have been completed, and in Christian et al., [2003] the geographical distribution of lightning and estimation of global flash rate were obtained.

Although OTD was a prototype sensor that had less sensitivity and navigational stability than its follow-on mission, the Lightning Imaging Sensor (LIS), data mining of the OTD dataset continues to provide valuable insight. For example, Boccippio et al. [2001] was able to obtain the climatological ratio of cloud flashes to ground flashes over the continental US by comparing OTD data with ground-based lightning observations obtained from the NLDN.

Since OTD made observations of a variety of lightning optical flash characteristics across the entire continental US, many additional OTD data analyses are desired. In this work, NLDN data is used to partition OTD flashes into ground and cloud flashes so that the optical characteristics of these two flash types can be compared. A fundamental question is asked: *Can the space-based optical measurements be used to discriminate ground flashes from cloud flashes?* This question is particularly important and relevant to the future GOES-R Geostationary Lightning Mapper (GLM); continuous knowledge of the *ratio* of cloud flashes to ground flashes derived from GLM data would provide a better understanding of thunderstorm dynamics, intensification, and evolution, and would improve the value-content of GLM data for severe weather warning.

2. METHODOLOGY

2.1 Identifying ground and cloud flashes

The approach for partitioning the OTD dataset into ground and cloud flashes is straightforward. First, the data for an OTD flash is read in. If the flash is associated with an instrument, platform, processing/algorithm, or external alert flag of any kind, or if the flash is not located over the continental US (i.e., longitude: -125° to -67°, latitude: 25° to 49°) it is thrown out. The flash is also thrown out if it does not pass routine Quality Assurance (QA) checks, or if it is suspected of being a noise event (i.e., the Thunderstorm Area Count, TAC, parameter is less than 140). Next, the NLDN dataset is scanned to see if the OTD flash is associated with an NLDN event. The OTD flash is assumed to be a ground flash if the NLDN event is within ± 0.5 s of the OTD flash time and within 50 km of the OTD flash centroid. The 50 km criterion is equivalent to the median OTD location error reported in Boccippio et al. [2000]; this location error is due primarily to satellite navigation errors and to a lesser extent to OTD pixel resolution limitations. The total time lag of the optical wavefront (due to cloud multiple scattering and the transit time from cloud-top to the OTD instrument) is accounted for when comparing all OTD/NLDN times. Since it is possible to have more than one NLDN event satisfy the (± 0.5 s, 50 km) criteria, the NLDN event closest in time to the occurrence of any optical group within the OTD flash is the NLDN event

* Correspondence to:

William Koshak, National Space Science & Technology Center, NASA Marshall Space Flight Center, Huntsville, AL 35805, USA. Email: william.koshak@nasa.gov.

that is associated with the OTD flash. If the OTD flash is associated with an NLDN event that has a positive peak current less than 15 kA, the flash is thrown out since it is suspected of being a cloud flash [Zajac and Rutledge, 2001; Cummins personal communication]. After applying all the filters mentioned above, a total of 45,913 ground flashes and 376,950 cloud flashes were obtained for the roughly 5 yr period.

2.2 Hypothesis tests

Inferences are made about the relative magnitudes of the *population* means of the flash parameters (e.g., radiance, area, ...) for ground and cloud flashes. A standard two distribution hypothesis test is applied on the population means to obtain rankings. For example, two distributions compared are the ground flash radiances (distribution 1) and the cloud flash radiances (distribution 2). The null hypothesis H_0 is written as: $\mu_1 \leq \mu_2$, where the μ s represent the population means. The alternative hypothesis H_1 for this case is stated as: "the population mean radiance of ground flashes is greater than the population mean radiance of cloud flashes; i.e., $\mu_1 > \mu_2$." The decision rule for rejecting the null hypothesis (i.e., for accepting H_1) with a 95% level of confidence is a right-tail test on a z-statistic [Aczel 1995]. In most cases, our results will actually end up exceeding the 95% confidence level.

3. RESULTS

3.1 Frequency distributions

The frequency distributions for OTD flash radiance and area are provided in Figure 1. The size, mean,

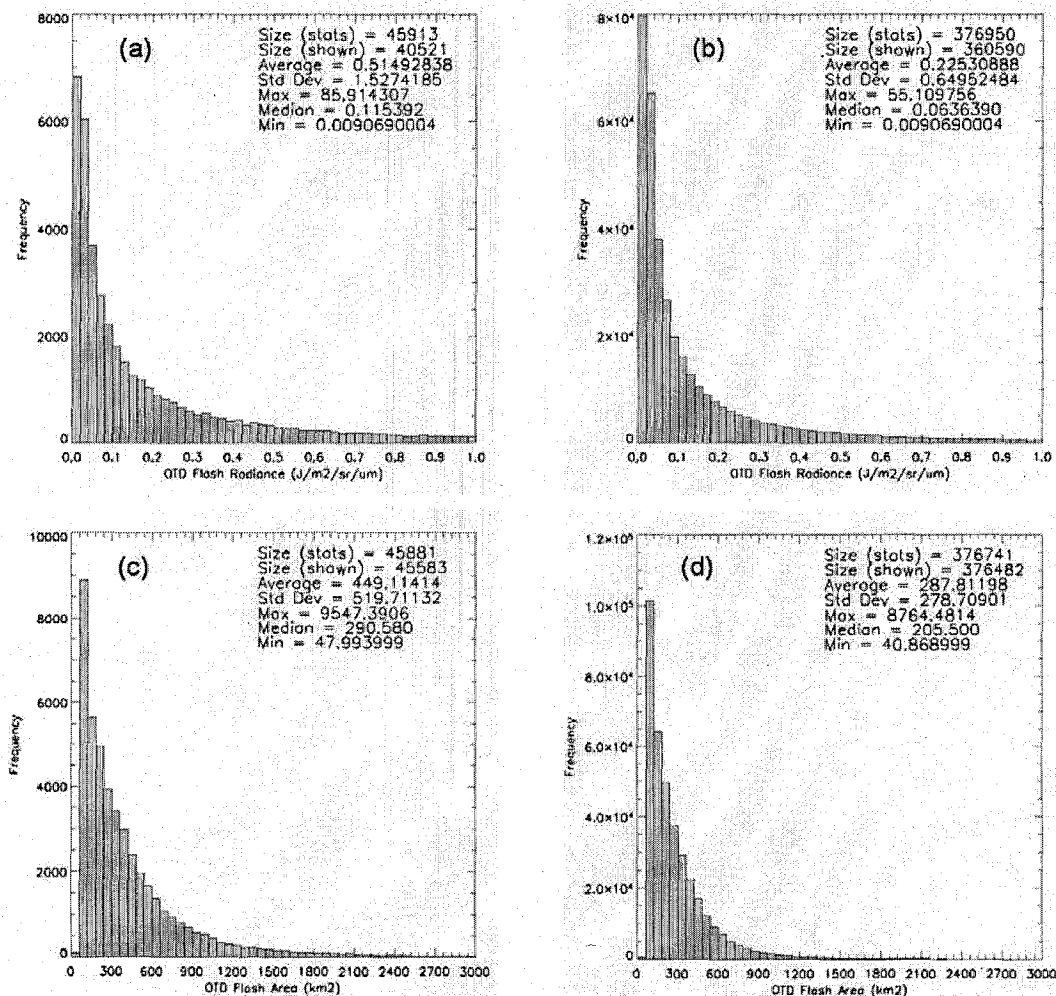


Figure 1. Frequency distributions of OTD (a) ground flash radiance (b) cloud flash radiance, and (c) ground flash area (d) cloud flash area.

standard deviation, median, max, and min of each distribution are provided in the upper right corner of each plot. The distributions for OTD flash duration, # groups, and # events were also obtained but are not provided here due to limited space; these distributions have exponential-decay shapes similar to those shown in Figure 1. Section 3.2 provides the means and standard deviations of all the distributions. Interestingly, both the mean and median values of the five parameters (radiance, area, duration, # groups, # events) are larger for ground flashes than for cloud flashes, except the median # groups was equivalent for the two flash types. The ratio of the average ground flash radiance to the average cloud flash radiance is 2.29, and the ratio of the median radiances is 1.81. The ratios of the means for the other 4 parameters range from 1.41 to 1.91.

3.2 Comparison of population means

Table 1 summarizes the hypothesis test results for comparisons between the ground and cloud flashes. Note that the null hypothesis (H_0) is rejected for each of the parameter comparisons. Moreover, the standard “p-value” [Aczel 1995] was nearly zero, so the confidence in rejecting each null hypothesis is nearly 100%. Physically, this means that one is highly confident that ground flashes are, optically speaking, more radiant, of greater areal extent, and longer lasting than cloud flashes on average. They also have more optical groups and events than cloud flashes, on average. In addition, positive polarity ground flashes were found to have larger average values than negative polarity ground flashes; this agrees with Koshak and Boccippio [2006] which used a different algorithm for matching OTD flashes with NLDN events, and a slightly larger analysis region.

Table 1. Hypothesis test results for comparisons of different mean parameters of ground and cloud flashes. [Note: the sample size for area and duration are smaller because the OTD processing algorithm sometimes zeroed-out areas and durations under certain conditions, and these cases were removed in this study so as not to adversely bias the statistics.]

Parameter	Size (CG)	Mean (CG)	Std Dev (CG)	Size (IC)	Mean (IC)	Std Dev (IC)	z	Result
Radiance (J/m ² /ster/μm)	45913	0.515	1.527	376950	0.225	0.650	40.25	Reject
Area (km ²)	45881	449.1	519.7	376741	287.8	278.7	65.35	Reject
Duration (sec)	34254	0.360	0.380	252946	0.255	0.257	49.63	Reject
# Groups	45913	7.597	12.28	376950	4.959	7.225	45.09	Reject
# Events	45913	17.86	36.64	376950	9.359	17.17	49.06	Reject

3.2 Bayesian inference

The derived frequency distributions and simple hypothesis tests obtained in this study suggest that it might be feasible to use space-based optical measurements to discriminate ground flashes from cloud flashes. It is proposed here that the general methods of Bayesian Inference [Gelman et al., 2004] are suitable for this task.

To employ the Bayesian approach, it is necessary to define required probability distributions. The frequency distribution given in Figure 1(a) can be converted into a probability distribution by dividing each bin size by the total sample size (= 45913). Because NLDN data was used to determine that the flashes in Figure 1(a) are ground flashes, the derived probability distribution is actually a conditional probability distribution. Similarly, Figure 1(b) can be converted into a conditional probability distribution. The two conditional probabilities can be written:

$$\begin{aligned} P(R|G) &= \text{probability of getting flash radiance } R \text{ given that the flash is a ground flash.} \quad [\text{derived from Fig 1(a)}] \\ P(R|C) &= \text{probability of getting flash radiance } R \text{ given that the flash is a cloud flash.} \quad [\text{derived from Fig 1(b)}] \end{aligned} \quad (1)$$

Here, G indicates ground flash, and C indicates cloud flash. However, it is the following “reverse” conditional probability that is ultimately desired,

$$P(G|R) = \text{probability that the flash is a ground flash given specific flash radiance measurement } R. \quad (2)$$

Bayes Theorem makes the connection between (1) and (2),

$$P(G|R) = \frac{P(R|G)P(G)}{[P(R|G)P(G) + P(R|C)(1 - P(G))]} \quad (3)$$

where $P(G)$ is the *prior probability*, i.e., the probability that a flash is a ground flash given no specific measurements of the flash. Note that $P(G|R)$ is referred to as the *posterior probability*, i.e., it is the probability after having considered the specific radiance measurement evidence R . If $n = \#$ ground flashes, and $N = \#$ cloud flashes, then $P(G) = n/(n + N) = 1/(1 + Z)$, where $Z = N/n$ is the ratio of cloud flashes to ground flashes in a typical thunderstorm. If one were totally ignorant of the value for Z , one could begin by asserting that $n = N$ so that $Z = 1$ and $P(G) = 0.5$. However, there have been many studies that give reasonable values for Z . Suppose one uses the continental US averaged value $Z = 2.94$ obtained in Boccippio et al. [2001]. This gives a value $P(G) = 0.254$. Now suppose that the space sensor measures radiance $R = 0.7$ for a specific flash. What

would be the probability that this flash is a ground flash?, i.e., what is the value of $P(G | R = 0.7)$? Using Figures 1(a) and 1(b), one obtains $P(R = 0.7 | G) \equiv 0.0041$ and $P(R = 0.7 | C) \equiv 0.0026$, so that (3) gives an answer $P(G | R = 0.7) = 0.349$. In other words, the radiance measurement increased the 25.4% prior probability to the posterior value of 34.9%, an increase of 9.5%.

The foregoing is just a simple demonstration of how one can extract information from one optical measurement, and its associated probability distributions, to upgrade a prior prediction. The Bayesian analysis can be generalized to include additional evidences, that is, one can consider the vector of space sensor optical measurements $V = (R, A, D \dots)$, where A = flash area, D = flash duration. The process for upgrading the prior probability given these several measurements is more complicated, but can be carried out using *Bayesian Networks* [Neapolitan, 2003]. Moreover, what has been demonstrated above is for the bulk case, that is, all the probabilities discussed have been for the continental US region as a whole. But, the spaced-based lightning sensors provide optical flash characteristics as a function of geographical location and this additional information should also be used. Hence, it is better to partition the continental US into $j=1, \dots, m$ sub-regions, and for each j^{th} sub-region specify the needed probabilities $\{P_j(R|G), P_j(A|G), P_j(D|G), \dots\}$. The prior probability $P_j(G)$ can be obtained from the geographical distribution of Z (the ratio of cloud flashes to ground flashes) obtained in Boccippio et al. [2001]. Bayesian network inference is then carried out on a sub-region by sub-region basis.

4. SUMMARY

OTD flashes occurring over the continental US have been partitioned into cloud and ground flashes using NLDN data. Large sample size frequency distributions for several OTD optical parameters (radiance, area, duration, # optical groups, and # of optical events) were obtained for each flash type, and basic hypothesis tests comparing the *population* means of these optical parameters were completed. The results indicate that there is a statistical significant difference between the cloud and ground flash optical parameters. Hence, it would be beneficial to exploit these (and possibly other) differences to discriminate flash type. Since several independent lightning observations provide a starting point for characterizing the climatological ratio of cloud flashes to ground flashes, it was suggested here that the techniques of Bayesian Inference are appropriately suited to ingest these prior predictions and then update them using the space-based, flash-specific optical measurements. The Bayesian analysis would provide a statistical statement about the probability that a given flash (occurring in a specific geographical region and having given optical characteristics) is a ground flash.

ACKNOWLEDGMENTS

The author thanks the staff of the Global Hydrology Resource Center (GHRC) and Dr. Dennis Boccippio for assistance in obtaining the OTD dataset, and Drs. Kenneth Cummins, E. Philip Krider, and Richard Blakeslee for their helpful discussions regarding small positive peak current events within the NLDN dataset.

REFERENCES

- Aczel, A. D., Statistics: Concepts and Applications, Chapter 7, *Richard D. Irwin Inc.*, 533 pp., 1995.
- Boccippio, D. J., W. Koshak, R. Blakeslee, K. Driscoll, D. Mach, D. Buechler, W. Boeck, H. Christian, S. Goodman, The Optical Transient Detector (OTD): instrument characteristics and cross-sensor validation, *J. Atmos. Oceanic Technol.*, 17, 441-458, 2000.
- Boccippio, D. J., K. L. Cummins, H. J. Christian, and S. J. Goodman, Combined satellite and surface-based estimation of the intracloud:cloud-to-ground lightning ratio over the continental United States, *Mon. Weather Rev.*, 129, 108-122, 2001.
- Boccippio, D. J., W. J. Koshak, R. J. Blakeslee, Performance assessment of the Optical Transient Detector and Lightning Imaging Sensor. Part I: predicted diurnal variability, *J. Atmos. Oceanic Technol.*, 19, 1318-1332, 2002.
- Christian, H. J., R. J. Blakeslee, D. J. Boccippio, W. L. Boeck, D. E. Buechler, K. T. Driscoll, S. J. Goodman, J. M. Hall, W. J. Koshak, D. M. Mach, M. F. Stewart, Global frequency and distribution of lightning as observed from space by the Optical Transient Detector, *J. Geophys. Res.*, 108, No. D1, ACL 4-1- ACL 4-15, 2003.
- Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin, Bayesian Data Analysis, 2nd Edition, *Chapman & Hall/CRC*, Boca Raton, 2004.
- Koshak, W. J., M. F. Stewart, H. J. Christian, J. W. Bergstrom, J. M. Hall, and R. J. Solakiewicz, Laboratory Calibration of the Optical Transient Detector and the Lightning Imaging Sensor, *J. Atmos. Oceanic Technol.*, 17, 905-915, 2000.
- Koshak, W. J., and D. J. Boccippio, Continental US ground flashes detected by OTD, *Lightning Imaging Sensor International Workshop*, Huntsville, AL, September 11-15, 2006.
- Neapolitan, R. E., Learning Bayesian Networks, Prentice Hall, 674 pp., 2003.
- Zajac, B. A., S. A. Rutledge, Cloud-to-ground lightning activity in the contiguous United States from 1995 to 1999, *Mon. Weather Rev.*, 129, 999-1019, 2001.